

DOCUMENT RESUME

ED 055 094

TM 000 381

AUTHOR

Medley, Donald M.

TITLE

Specifications for a New Teacher Examination: A First Approximation.

INSTITUTION

Educational Testing Service, Princeton, N.J.

PUB DATE

Sep 70

NOTE

89p.

EDRS PRICE

MF-\$0.65 HC-\$3.29

DESCRIPTORS

Behavioral Objectives; Biographical Inventories; *Effective Teaching; Film Study; *Predictive Ability (Testing); Scoring; Situational Tests; Teacher Characteristics; *Teacher Evaluation; Teacher Qualifications; Teacher Selection; *Teaching Skills; Teaching Styles; *Test Construction; Test Interpretation

IDENTIFIERS

*Educational Testing Service

ABSTRACT

Suggestions are offered for constructing a test which would predict how a teacher would teach, thus providing school personnel officers, teacher certification agencies, and other clients with an index of the degree to which a candidate has mastered the content of teacher education. Three tasks in constructing this type of teacher examination are emphasized: (1) Defining the domain of behaviors contributing to success in teaching that the test must sample (Chapter II); (2) Specifying the domains to be covered by the modules of which the test is composed (Chapter III); and (3) Constructing items to put into the modules (Chapter IV). Although all of the suggestions offered for the new teacher examinations involve innovation in varying degrees, it is emphasized in Chapter I that a significant change in scoring and interpreting the test is needed. A reporting on cognitive factors in teaching style is appended. (AG)

ED0 55094

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION

PR-70-14

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

SPECIFICATIONS FOR A NEW TEACHER EXAMINATION:

A FIRST APPROXIMATION

Donald M. Medley



September 1970

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

I. BY WAY OF INTRODUCTION: A MODEST PROPOSAL

Perhaps a good way to begin this report is by examining the basic function which a teacher examination service should attempt to perform for the educational community. Currently, the function performed by the National Teacher Examinations is probably best described as follows: to provide school personnel officers, teacher certification agencies, and other clients with an index of the degree to which a candidate has mastered the content of teacher education. To a lesser extent, perhaps, it also seeks to provide some diagnostic information--that is, some indication in which parts of that content the candidate may be most or least well prepared; this function, however, is not very well developed.

In thinking about a teacher examination one's first thought is that such an examination should provide clients with an index that is predictive of success in teaching. Such an index would indeed be useful, and the idea of building such an instrument has great appeal. But success in teaching is dependent on so many variables not accessible to measurement at the time when most candidates need the index--that is, at the point when they apply for their first teaching positions--as to make the construction of such an instrument much more difficult than it would appear at first glance.

Moreover, there is neither sufficient understanding of the nature of teaching aptitude to provide a base for constructing such a test, nor any one defensible criterion against which it could be developed empirically. Whatever makes a teacher successful is highly specific to little-known elements in the situation in which he is to succeed; and the task of deciding which teacher is a success and which is not is far from simple.

Another way of conceptualizing a teacher examination service is as an information service designed to provide clients interested in teacher selection, retention, promotion, certification, etc., with information relevant to these interests in a form which would facilitate their using it in whatever way they see fit. Among the information supplied, estimates of a candidate's mastery of various areas of content included in teacher education might well be included, but such data need neither be the only data supplied nor need there be any prescription from ETS as to how the various areas should be weighted to form a composite score to be used in selecting, retaining, promoting, or certifying teachers. Deciding what a teacher should know (or what other characteristics he should possess) before he is permitted to teach in a certain locale is the responsibility of some local official or agency, not of ETS. As far as factual items go, ETS has a responsibility to key items correctly and should have the competence to do so. But it has neither the responsibility nor (with or without competent outside help) the competence to weight items to produce a composite that will predict teaching success in that situation.

If the new examination service is to offer anything more in the way of information than the inventories of content knowledge described above, the situation changes. If the examination moves in the direction of classroom performance, if it, for example, asks a candidate how he would behave in a specified classroom situation (or how he has behaved in the past), what basis does ETS have for saying what he should do (or should have done)? What competence has an examination service to key such items? Might it not be that in certain situations, in certain

school systems, in communities with certain values, one way of proceeding would be considered optimal while in another system or community an entirely different course of action would be considered successful?

It might be possible to limit problems used in the examination to ones on which there is consensus, so that there would be one generally recognized and accessible right answer, and only one, to each item.

Unfortunately, the teacher on the job has no such control over the problems he must face; and the ones about which there is no consensus tend to be the most difficult ones. A test based only on problems which have solutions about which everyone can agree is likely neither to be representative of teaching problems in general nor to predict a teacher's ability to solve them.

All of which seems to strengthen the argument for placing the responsibility for saying how a teacher examination is to be scored somewhere closer to where a candidate is going to teach than here at ETS. When a client--a school personnel officer, for example--wishes to use the examination, the first thing he must do is specify how he wants the various parts of the examination weighted; he must indicate in this way what he wants his teachers to know, how he wants them to deal with classroom problems, what patterns of experience and/or attitudes he wants them to possess, and so on.

The simplest way to do this might be to have the client sit down and work through the test himself. On those items which are factual in nature he could indicate whether that item of knowledge was important or not. On problem items, he should indicate how he would prefer to have his teachers respond to it, as well as how much weight should be attached to the answer.

This task should be streamlined by grouping knowledge items of like type into clusters or modules, and assigning weights to each module. Responses to problem items should also be grouped according to overall strategies or approaches to teaching which they reflect, and the client might then assess the strategies as to their relative desirabilities for his own situation.

This approach would enable each user of the test to specify a key which, when applied to a set of test performances on file at ETS would yield a single composite score on each performance whose magnitude would be a direct function of the match between that candidate's test performance and the ideal specified by the user. On request the user could call for a set of scores on some defined group of candidates, or for a list of (say) the 50 available candidates best fitted to his specifications.

If the examination record included biographical data and information about what kind of teaching position each candidate would accept, this latter approach would seem to be particularly useful. It would, of course, require that candidate's performances be stored at a level of detail not necessary with the present examination and would probably require more computer time than is presently used, with a resultant increase in cost to the client. On the other hand, a much more individualized and potentially useful service would be provided.

If this approach was adopted, a service to the candidate himself not presently available could also be offered--that is, a diagnostic profile of his performance on the examination (rather than a total score) which would be immediately useful in indicating areas where he needed

further study or training, and potentially useful in indicating the kind of position in which his pattern of abilities, etc., best fit him to succeed.

This approach could, of course, be followed without any major change in the present teacher examination. All that would be involved would be a change in the way items would be selected; the criterion for assessing the discriminating power of an item would be its correlation with the subtest or module to which it belonged, rather than with total test score. It is to be expected that within modules internal consistency would be maximized but that between-modules correlations might drop, possibly reducing reliability of total scores somewhat. But of what intrinsic value is high internal consistency in a test including such heterogeneous elements as ability to solve algebra equations and ability to recognize emotional upset of pupils? Movement in this direction would be an important, and relatively simple, first step to take in improving the Common Examinations.

It would also be consistent with this step to abolish total or composite scores, requiring only profiles of module scores to users. While it is true that none of the subtests or modules have any greater or less validity than any of the others as overall predictors of teacher competence, it can also be shown that they do measure different abilities. Maybe for some applications it would be better to ignore some, or pay more attention to others. Clients have a right to that much flexibility. It has also been shown that there is a wide range of internal consistencies in different subtests, which suggests that some could be shortened considerably, particularly if only the most discriminating items (against subtest score, not total score) were retained.

But the potential of this modular approach to test use cannot begin to be exploited until the content and form of the test items are changed--until an examination is constructed which will go beyond measuring a candidate's mastery of the cognitive content of teacher education and predict how successful he will be on the job. The trouble of figuring out weights on modules would be worth taking if the modules predicted performance!

There is presumptive evidence that the task of constructing a test that will predict how well a teacher will perform as a teacher before he tries is impossible.

Item: Nobody understands what makes a teacher effective. Then how is it possible to select and construct items for a test of teacher competence?

Item: There is no satisfactory criterion of teacher success against which such a test could be validated if it were built.

Item: Teacher effectiveness is a complex, perhaps even idiosyncratic trait: different teachers in different situations achieve comparable effects in different ways. Then is it possible even to conceive of a single test which would be valid for different teachers in different settings?

Looking at these three items is a discouraging experience for a test constructor. Item one seems to rule out an approach through content validity. Item two cuts off an empirical approach. Item three implies that an internal consistency approach will not work. What is left?

The solution is to use a bit of each. Begin by selecting items on the basis of content--items that seem likely to relate to teacher

competence. Caution: do not require universal or even general agreement among "experts." Be hospitable: include any characteristic that is seriously defensible as likely to relate to teacher competence. Think of items not as having demonstrated validity but as representing hypotheses about validity to be tested later.

Then use a bit of an internal consistency approach to organize items into internally consistent subsets or "modules," letting inter-module correlations fall where they may. This should produce a test that is interpretable, not as yielding one meaningful score but several distinct ones, each meaning what it means in a content validity sense, but without any empirically demonstrated relationship to teacher competence, or performance.

Finally, pass the validation problem on to the client: ask not how valid the test is but where it is valid. Work with each user, first to help him define a composite of the meaningful modules which operationalizes his best guess as to what makes a teacher successful in a specific situation. Who is closer to the situation than he? If there is anyone closer, invite him to help. Work also with the user after he has used the test to find out how well it is working and how the composite might be altered to improve its fit--that is, its validity. Pay no attention to what works somewhere else, but concentrate on getting a composite that works here. Grass roots validity, that's the idea.

If anyone wants to study the validity problem in a broader context (such as a doctoral student somewhere) the modular examination should suit him very well, and he should be encouraged. It may well be that some parts of the examination will turn out to relate to teacher competence in any

setting--but until there is evidence to support this assumption it should not be depended on. Who needs it?

This is the task and the general approach that must be used in accomplishing it. In the pages to follow the reader will find incomplete and highly imperfect first passes at three tasks that need to be done. The first, defining the domain of behaviors that the test must sample, is the concern of Chapter II. The second, specifying the domains to be covered by the modules of which the test is composed, is assailed in Chapter III. And the third, that of constructing items to put into the modules, is treated in Chapter IV.

II. THE DOMAIN OF TEACHER COMPETENCE

One of the greatest problems connected with the specification of that domain of behaviors which may properly be called teacher competence has always been that of weighting the behaviors proposed for inclusion. Especially important has been the problem of deciding which items should get zero weight--that is, be eliminated--and which should get non-zero weights--that is, be included. Decisions about these problems require judgments as to the relative importance to success in teaching of such things as knowledge of quadratic equations, familiarity with the writings of Sir Walter Scott, understanding of the dynamics of small groups, ability to identify appropriate strategies for dealing with learning difficulties, etc.--judgments by no means easy to make.

This problem becomes particularly acute when it is faced by an independent service organization like ETS, which has no constituency to represent and no responsibility to any community, school system, or training institution. Since ETS' own set of priorities--if it has one--has no standing whatever, and since there is no justification for adopting the values of any single group and foisting them on all the others, basing weights on a consensus would seem the natural solution. This has two drawbacks.

For one thing, a consensus in this area tends toward a lukewarm, wishy-washy definition that threatens to eliminate any candidate likely to be more exciting or even very different from the majority--as indeed any really outstanding teacher is likely to be. And for another, a consensus threatens to be really satisfactory to no one.

The way around this dilemma recommended here (and adopted as a working plan) is to eliminate from the domain no element that is seriously defended by any group--to define the domain in effect as including any behavior that has a chance of being related to competence, and sampling the domain as widely as possible when building the examination. Concerned clients can then define weights to suit their own conditions, which weights can be used to generate scores tailored to those clients' needs (which may differ substantially from scores reported to other clients).

Bearing this in mind, the reader should realize that the inclusion of an item or area in the domain does not imply an endorsement of it as an important element in teacher competence, it merely proposes a hypothesis that that item may be important somewhere--even though some effort will be expended in defending each area included.

The behavioral characteristics that have been proposed as ones necessary or important to success in teaching can be conveniently organized in terms of areas or clusters of behaviors which might be referred to as roles the teacher is expected to perform successfully. A good teacher, it has been argued, must be a well-educated or cultured person. He must be a sound scholar in the subject he teaches. He must be a skilled practitioner of the craft of teaching. We must have the attitudes, values, and knowledges of a professional. He must be mentally healthy--emotionally mature and stable. These are the five roles that will be used here.

Serious consideration of definitions of teacher competence have at least two immediate effects. One's first reaction is to reflect that it is no wonder there are so few good teachers around if this is what it takes; and the second is to wish that popes, presidents, and premieres

were required to qualify as good teachers, and to speculate on what a wonderful world this could be if they did.

The domain of behavior that will be defined must be regarded, then, as defining an ideal which few teachers will approach closely. To put it another way, it defines a lot of desirable characteristics, some of which one candidate will possess, some of which another will possess. The client interested in identifying teachers able to perform successfully in a particular situation must somehow decide which, characteristics are most important in that situation, and try to identify in a pool of available candidates those who fit his specifications best.

This does not suggest any need (or particular use) for a test which samples the entire domain and estimates what proportion of all of characteristics in it a candidate possesses. It suggests a "test" which describes which characteristics each candidate has and which he lacks, so that selection may be based on those abilities most nearly relevant to a specific situation or position.

It is in this spirit that the following set of characteristics is presented, and also with the frank admission that it may be, and indeed almost certainly must be, incomplete.

1. The General Culture Area: The Teacher As An Educated Man

This part of the domain seems to be of primary concern to parents and laymen, although its importance is widely acknowledged. Bestor (1953, p. 20), for instance, argues that no one can communicate a liberal education to others who does not possess one himself. Certainly a teacher who is unaware of the cultural heritage cannot relate the everyday events in his classroom to it. A teacher who himself misuses the English language as he teaches

some other subject can do more damage to his students in an hour than the English teacher can undo in a week. Some parents put their children in private schools at considerable expense rather than in public schools, under the assumption that more culture will rub off from the private school teachers (who tend to be more cultured) than can be taught by the public school teachers (who tend to be more better trained in teaching methods). There is little reason to doubt that children imitate adults whom they admire, or that well educated teachers can enrich their instruction by relating it to other subjects, countries, etc.

An element in this picture that has come into focus relatively recently is the better understanding and communication between teacher and pupil that is possible when they share a common cultural heritage and background, manifested in a host of ways--notably in the intelligibility of the vocabulary of the teacher and in the degree to which implied purposes are understood by both parties. This makes it appear that the qualified teacher needs to be well grounded not only in what is usually referred to as the common culture of the society but also in the sub-cultures of his pupils as well, so that he can assist his pupils in understanding and adapting to the former without losing sight of the unique values of the latter.

All of these things ought, of course, to be accompanied by the enlightened curiosity and the habits and skills of inquiry and appeal to reason that are recognized as the principal effects of a liberal education; as well as a concern for humanity, commitment to freedom of inquiry, and faith in the continuing search for truth as a means of improving the human lot. Among the behaviors one would expect to characterize the competent

teacher which would fall into this general area one might identify the following components:

(a) General information. The teacher can and does draw for illustrative purposes on a wide knowledge of many subjects--mathematics, the sciences, literature, music, art, history, politics, etc.--both current and traditional, and in relationship both to the "common" cultural heritage of mankind and that which is unique to the culture of his pupils.

(b) Inquiry habits and skills. The teacher locates quickly and as a matter of course information relevant to new topics he encounters. He customarily reads books and magazines related to a wide range of interests, some continuing and others merely timely.

(c) Humane values. The teacher displays concern for and actively works to remedy some of the ills of the world. His behavior reveals a belief in freedom of inquiry and the democratic process, and their manifestation in the form of cultural diversity.

2. The Subject-Matter Component: The Teacher As A Scholar

The impetus to require that the teacher be a scholar in at least one subject may be primarily identified with the learned professions, although the requirement is much more widely endorsed. A typical manifestation is the reaction to the oft-repeated tale about how Albert Einstein was not eligible for a certificate to teach mathematics in the public schools of New Jersey. Most people's reaction to this anecdote suggests that they may regard mastery of one's subject as not only a necessary but also a sufficient condition for teacher competence! Other groups tend to minimize the importance of scholarship in a field--especially for such groups as elementary teachers.

Part of the case for scholarship was succinctly stated (and confused with that for a liberal education) by Bestor (1953):

Liberal education, in other words, is essentially the communication of intellectual power. That it cannot be communicated by someone who does not possess it--by a teacher who is not also a scholar--is self-evident.

Those who define teaching as something else besides transferring what is in the teacher's head into the pupil's might quarrel with this argument, but even they tend to value mastery of the discipline taught for a reason (also acknowledged by Bestor) perhaps best defended by Bruner (1960, pp. 17 ff), that of a subject is to be taught as a subject rather than as a collection of facts, the structure of the subject or discipline must form the main content, not the facts. Unless a teacher has himself been well grounded in a subject, unless he is at least an embryonic scholar, he cannot either communicate to students or help them discover the structure of a subject.

Much the same argument holds for the subject as a discipline. Each scholarly discipline incorporates its own means of inquiry, its own method for developing new knowledge; and no one who does not understand this methodology has any real understanding of the subject or can lead students to such an understanding.

Almost equally important is another characteristic of the true scholar--his devotion to, his enthusiasm for his discipline; his belief that in the application of his discipline (whatever it may be) lies the key to a better future for mankind. This enthusiasm should at least engender in his students a respect for the discipline, and at its best may fire them with a lifelong interest in it.

One aspect of this matter that is sometimes overlooked is that the subject-matter teacher should also understand his subject (content, structure,

and discipline) in a way that relates to learning it. This is a notion that is difficult to explain clearly, but has to do with knowing which concepts are the difficult ones, which ones illuminate others, and so on. Such knowledge tends to be possessed by those who understand a subject best.

Four components of teacher competence in the area have been identified, then, which may be summarized as follows:

- (a) Subject-Matter Content. The teacher draws on an extensive knowledge of facts, terms, concepts of his subject.
- (b) Subject-Matter Structure. The teacher relates content to the structure of his subject, and helps students to become aware both of the structure itself and of where content fits in.
- (c) Subject-Matter Discipline. The teacher uses the methods of inquiry of the subject himself.
- (d) Enthusiasm for Subject. The teacher communicates a feeling of excitement about his subject to his pupils.

3. Instructional Skill: The Teacher as Teacher

It should not seem necessary to defend the importance of instructional abilities to teacher competence, but there are some who dispute the very existence of such abilities, at least as something that needs to be studied. Whether these abilities are innate and unteachable or not, and whether one's concept of the function of a teacher is that of an information given or of a promoter of independent inquiry, there are abilities a teacher can find useful which ought to be measured.

First of all, the teacher needs to possess a repertory of skills he can use, analogous to the kit of tools a craftsman needs. Whatever the behaviors are that affect pupil learning, the teacher needs to be able to exhibit them when he needs them.

Next, he needs to be able to understand pupil behaviors well enough to be able to relate his own behaviors to them--he needs to be able to judge how well pupils understand, when they are puzzled, when they are bored. And as he gets these cues from his observations of pupils he needs to be able to select from his repertory those techniques which will accomplish his purpose.

The teacher needs to be able to create and maintain a classroom environment favorable to learning. This may mean no more than keeping things quiet or it may mean creating a psychological climate in which such activities as reading, listening, and working on arithmetic problems are seen as desirable, fun things to do; and such activities as making loud noises, running around, and so on are not approved of by the group.

The teacher needs to be able to size up a group or an individual pupil, his environment, his interests, his capabilities, and design a set of objectives and a sequence of learning activities suited to them or him; he needs to be able to design curricula as well as implementing them.

These characteristics can be summarized under five components of teacher competence as follows:

- (a) Instructional Skills. The teacher demonstrates at will a large number of skills and techniques for facilitating learning.
- (b) Sensitivity. The teacher correctly interprets behaviors of students.
- (c) Flexibility. The teacher adapts his use of techniques and materials to the behavior and needs of his pupils.
- (d) Planning. The teacher identifies appropriate goals and designs learning experiences likely to result in their attainment.
- (e) Style. The teacher exhibits consistent behavior patterns which produce a classroom environment favorable to learning.

4. Professionalism: The Teacher as Professional

In addition to the instructional skills included in the last section, there are a number of other professional skills, knowledges, and attitudes a teacher would seem to need to become a fully competent teacher. These skills may be characterized as not involving pupils directly but nevertheless seen as important to success in teaching.

A teacher needs to know how to work with other members of the professional staff, with parents, and with the community at large. He needs to have a clear grasp and understanding of the place of the school in society as a whole and in the particular segment of society in which it is located, so that his decisions about what to teach--and how to teach it--maybe wise ones. He needs to be able to read professional literature--including statistically sophisticated research studies--and relate the results to his own situation. He must be aware of the new curricula, methods, materials, and technology that are constantly being developed, and able to evaluate them critically. He needs to know how to experiment with his own teaching--how to analyze his own behavior and its effects on pupils, and how to assess the potential of innovation for his own classroom, and above all to have a permanent interest in self-improvement.

Some components of teacher competence which might be identified in this area are listed below:

(a) Professional Awareness. The teacher reads and keeps informed in other ways about the issues and innovations which concern the profession at any given time, as well as those that have been continuously studied and discussed, and how they relate to his own school, community, classroom. He evaluates what he reads critically, whether it be a research report, a magazine article, a book or whatever, and relates it to his own problems.

(b) Self-Improvement Skills. The teacher is competent to try out new techniques and materials and evaluates their effects. He analyzes his own behavior, diagnosis difficulties, and takes steps to remedy them. He is continually questioning what he does and never fully satisfied with his present level of competence.

(c) Interpersonal Skills. The teacher works productively with professional colleagues, supervisors, and with parents and other members of the community. He uses the facilities and support mechanisms of the school and the resources of the community effectively and with understanding. His curricular and procedural plans forward the goals of the school, the community, and the society.

(d) Professional Attitudes. The teacher demonstrates the social motivation, the high ethical standards, and the commitment to the educational process as a vehicle for improving the human lot appropriate to his calling.

5. Personality Integration--The Teacher As A Mature Human Being

Some students of the teaching process have stressed the development of acceptance and understanding of one's self as a crucial step in becoming a competent teacher. Only the teacher who has learned to understand and accept his own strengths and weaknesses as they exist is free to deal with problems not related to his own concerns (cf. Fuller, 1969). Only a teacher who has matured in this sense of the term is able to view the problems he encounters in his teaching with the professional detachment necessary for effective coping. Only the teacher who has learned to accept himself as he is can accept the scrutiny and criticism of himself, his supervisor, and his peers.

III. MODULES FOR A TEST OF TEACHER COMPETENCE

The proposed new examination will be made up of a series of elements called modules. Each module consists of a relatively small number of items or item responses keyed to reflect a certain characteristic which has been identified as a potential predictor of classroom performance. The module should have very high internal consistency, and correlation with total score on the module should be the criterion used in item selection. Correlation with total score on the examination or on any portion outside the module is entirely irrelevant to the question of item selection. The ideal distribution of scores on a module would be not normal but bimodal--tending to divide candidate into two distinct groups. If an outside criterion measure of the characteristic the module is supposed to measure is available for item selection purposes it might be a better basis of item selection than the total score on the module, or it might not, depending on the nature of the characteristic. In many instances, such an outside criterion would be too unreliable to be useful. What is being sought at this point is interpretability of scores on the module so that its true nature can be communicated to test users clearly enough for them to decide how the module should be weighted for selecting teachers for their situations, and so that candidates themselves can interpret scores for self diagnosis.

Intercorrelations between different modules are not to be used as a basis of item selection but rather will be objects of study. It is to be expected that they will be positive, since all have been defined as part of the same general trait, that is, teacher competence; but it is conceivable that some might show negative relationships to others.

The obvious procedure in constructing a module would be the usual one of building several times as many items as are needed, trying them out, doing an item analysis, and selecting those which have highest discriminating power. It might be worthwhile to use an iterative process, eliminating the least discriminating items and then recalculating item--item-module correlations against those left.

In the pages to follow specific suggestions will be made as to what modules might be constructed, and what the items in them might measure. This list should be regarded as a first approximation only, representing as it does the often arbitrary decisions of one judge. Panels of qualified and concerned persons should be convened to review these modules, to define new ones to fill important gaps, and to reorder the ones here as necessary.

In setting about the task of defining this first set, heavy use has been made of the 10 Model Elementary Teacher Education Proposals constructed in response to a Request for Proposals issued by the Bureau of Educational Research at the United States Office of Education in 1967. Use was made of, and acknowledgement of invaluable assistance is hereby made not only to the nine Phase I reports funded by OE (prepared at the following universities: Columbia, Florida State, Georgia, Massachusetts, Michigan State, Pittsburgh, and Syracuse, and by consortia in Ohio and Oregon), as well as to one that was not (the Wisconsin Model). Use was also made of a curriculum design developed at Fordham (Rivlin & Robinson, 1968).

The procedure followed was to go through the lists of performance criteria included in each proposal, drawing out those which described behaviors which might conceivably be accessible in the testing situation envisaged. Several hundred of these were identified. Next they were sorted

under the five major areas described in Chapter I, duplications were eliminated; and finally the criteria were grouped under sub-categories and "components" suggested by the nature of the criteria themselves.

The vast majority of the criteria fall into the third and fourth areas, the two that have most directly to do with on-the-job performance. So few were found on the other three areas that no attempt has been made to base a set of modules on them. The two areas into which the bulk of the performance criteria fall, the instructional and the professional, are based entirely on the criteria. The project staff felt no competence to describe what elements should be included or excluded (except on the basis of accessibility to measurement), so the resulting set of modules should reflect the particular biases which characterized these programs as a group. The fact that general education, subject-matter, and personal maturity are virtually absent, (as well as certain gaps in the area of professionalism) from the list of criteria clearly reflects the mandate given the module builders to focus on classroom performance rather than on knowledge as such.

For this reason, the three "neglected" areas to follow will perforce be less thorough and less firmly grounded than that of the "performance" areas.

It would seem that these latter two areas are the ones which should receive highest priority in the development of the new examination for two reasons. One reason is that this part of the examination is likely to produce a large portion of the validity of the test as a predictor of competence than the rest. And the other reason is that this is the part of the test that should correlate most closely with how a teacher behaves

in the classroom, and thus be most useful in the long period that must pass before competence is well enough understood to be predicted successfully.

One other point should be mentioned. The modules are defined here in terms of the behaviors they should measure or predict, not in terms of the procedures to be used in measuring (or predicting) them. The modules should, and it is hoped will, suggest strongly what kind of items should be written--indeed, this is their main purpose--but they do not and should not specify item format.

Area 1. Modules In The Cultural Area

As has been mentioned, the Model Programs have been of relatively little use in developing modules for this portion of the examination. It appears that that large and vocal group who advocate a strong liberal education as perhaps the most important element in a teacher's preparation tends to be made up of people who have little sympathy with or understanding of the concept of performance criteria, and that the group who are convinced of the importance of performance criteria in teacher education tend to value general culture less--perhaps in part, because it is difficult to define performance criteria related to teaching behavior which involve the area.

But the nature of the new examination makes it possible to include these areas without the need for showing that performance related to general culture is related to performance on the job at all; the existence of such a relationship is a problem for investigation after the test has been built.

Components in three cognitive sub-areas should be tapped, which may be referred to as Knowledge, Skills and Habits, and Minority Culture. In addition, a case might be made for a non-cognitive module related to values and attitudes.

Knowledge Component. This component should indicate the amount of knowledge a candidate possesses of each major content area represented in the school curriculum. As a start, the seven areas included in the present Common Examinations might be used.

- Mod 1. Science Knowledge
- Mod 2. Mathematics Knowledge
- Mod 3. Mechanics of English
- Mod 4. Effectiveness of Expression
- Mod 5. Literary Acquaintance
- Mod 6. Fine Arts Knowledge
- Mod 7. Social Studies Knowledge

A study of the internal structure of the examination recently completed (see Appendix) indicates that each of these subsets of items on the present examination does measure something different from the others: if an internal analysis were made of each subtest, calculating the item subtest correlations for each item, it is likely that the number of items on each could be reduced and that the smaller number of highly discriminating items remaining would still discriminate as well as the present sets--or nearly so. This could reduce the total testing time on this component to an amount commensurate with its importance in the entire test.

Skill and Habit Component. The notion of a liberal education, from which a good deal of justification for modules in this area comes, includes much more than the mastery of a certain domain of knowledge which the first seven modules are designed to assess. No help was found in defining modules in the Model Programs, as has already been pointed out; and there is no other source with any standing from which a definition of the area can be derived.

The following areas are suggested as illustrative of the kind of modules that might be expected to emerge.

Mod 8. Critical Thinking

Mod 9. Habit of "Serious" Reading

Mod 10. Inquiry Skills

These modules are deliberately left undefined for the present, since they are meant only to be suggestive.

Minority Component. It has become more and more apparent that this area of general culture should include modules which assess a teacher's familiarity with minority cultures. The principle rationale is that such knowledge lays a better base for a teacher's understanding of pupils from these minority groups, and may therefore enhance his effectiveness in teaching them. It is also possible to defend such modules as measures a teacher's interest in working with minority group pupils, on the grounds that anyone who is really interested in something will know something about it.

Again there is no basis for specifying all of these subcultures, but following are some which might be included.

Mod 11. Knowledge of Urban Black Culture

Mod 12. Knowledge of Chicano Culture

Mod 13. Knowledge of Appalachian Culture

These 13 modules suggest the kinds of modules which should ultimately be constructed in the area of general culture. As has been pointed out, they are far from definitive, and a much more careful analysis should be made before modules are actually constructed.

Area 2. Modules In The Subject-Matter Area

Because there are so many subject-matter areas and other areas of specialization within the teaching profession, it is altogether impractical to attempt to define modules for this important part of the domain of teacher competence. For the present it is recommended that the present Teaching Area Examinations serve as modules in this area as they are. When all of the other areas have been "modulated" it may be time to turn to this one and rework it, at least to a point where scores on the four components identified in Chapter II can be scored. It is also recommended that the present policy of letting the candidate choose which area tests he wishes to take also be continued.

Area 3. Modules In The Teaching Area

As was pointed out in the beginning of this chapter, modules in this area were derived inductively from specific performance criteria included in one or another of the Model Elementary Teacher Education programs; some of these criteria are listed with each module to generate ideas for test items in that module.

Component 1. Instructional Skills. This component has to do with the teacher's command of a repertory of professional skills, methods, etc.

Mod 14. Ability to apply principles of learning in the classroom

Identify behaviors typifying major principles of learning theory

Identify ways he could use cueing to provide success experiences
for less able students

Recognize the use of reinforcement in both simulated and real
interactions with pupils

Identify ways in which different forms of reinforcement-physical,
social, intrinsic, extrinsic can be used

Identify specific ways of rewarding pupils

Mod 15. Skill in asking questions

Discriminate among factual, conceptual, and value questions

Identify probing questions that assist in finding out about the
pupil's information processing system

Choose a series of questions which lead students to master each
part of a skill or each element of a concept, as the case may be

Identify questions which call for inferential thinking, evaluative
thinking, building of generalizations, or identifying criteria

Identify questions which challenge students to make hypotheses,
project themselves into historical situations, guess at
solutions to unsolved problems, dream up new ideas

Mod 16. Content presentation skills

Identify technical skills related to content presentation--e.g.,
set induction, closure, probing, planned repetition, use of
examples

Recognize and distinguish such processes as defining, describing,
designating, stating, evaluating, classifying, conditional
inferring, explaining

Discriminate among various types of explanations--sequential,
mechanical, procedural, normative, causal, teleological

Component 2. Sensitivity. This component seeks to measure the candidate's
ability to secure accurate feedback from pupils--that is, to understand them
and their behaviors.

Mod 17. Sensitivity to pupil behaviors

Recognize the key dimensions of attending behavior: eye contact,
physical attention, and verbal thought and behavior

Recognize behaviors of pupils that indicate concept learning,
principle learning, problem solving

Recognize cues (facial expression, body postures) which indicate
interest level

Recognize cues to pupil's developmental level, conceptual style,
and frame of reference

Mod 18. Knowledge of behavioral manifestations of human growth and
development

Identify and recognize the phases through which a child must
progress to achieve healthy personal development

Identify similarities and differences in growth patterning of
males and females

Categorize preschool children according to physical development

Identify visual, auditory, and cognitive perceptual development
levels of children

Mod 19. Awareness of cultural differences in pupil behaviors

Identify the social and cultural determinants of behavior in
classrooms

Evaluate teaching strategies in terms of the character of the
community

Recognize sociological variables which affect instruction

Identify cross-cultural differences within the urban setting

Mod 20. Use of tests and measurements

Identify, compare, contrast measurement devices used by teachers

(with respect to ease of administration, scoring, etc.)

Evaluate research relevant to a measuring instrument

Interpret scores on the instrument

Judge appropriateness of an instrument to a stated purpose

Identify desirable test characteristics

Component 3. Flexibility. This component is intended to evaluate a teacher's ability to adapt his behavior to changing purposes and conditions.

Mod 21. Ability to adapt to individual differences

Identify ways of modifying teacher behavior in response to specific
pupil behaviors

Recognize differences between students and between groups and

identify strategies and style of teaching appropriate to each

Judge competency in matching instruction to strengths and weaknesses
unique to each child

Discriminate learners as to cognitive orientation and evaluate the
teaching strategy accordingly

Evaluate strategies based on an analysis of social needs of pupils
in a given situation

Mod 22. Teaching judgment

Evaluate maneuvers designed to induce productive thinking

(generate hypotheses, synthesize information, build
generalizations, etc.)

Evaluate maneuvers intended to induce mastery of content or skills

(demonstration, recitation, programmed techniques, etc.)

Evaluate maneuvers intended to induce self-direction (role playing, discussion, counseling, etc.)

Evaluate maneuvers designed to structure activities (initiating tight organizational procedures)

Mod 23. Knowledge of teaching strategies

Relate teaching strategies to theoretical positions on learning, philosophical stances, and ways of organizing and analyzing and disciplines

Recognize a wide range of teaching strategies

Analyze teaching strategies in terms of amount of external structure, task complexity, provisions for teaching students their roles.

Mod 24. Skill in developing independent learners

Identify ways to help a pupil assess his own strengths and weaknesses

Identify ways to aid individuals and groups to assess their progress toward defined goals

Evaluate attempts to encourage pupils to seek knowledge for themselves

Identify ways to help pupil structure his own goals and activities

Evaluate efforts to turn attention of pupils toward analysis of their own concepts and strategies

Recognize attempts to encourage children to contribute to the planning of learning experiences

Identify ways of eliciting opinions and suggestions of students

Mod 25. Ability to develop inquiry skills

Identify situations in which students are likely to raise hypotheses or suggest alternative solutions

Recognize attempts to involve pupils in intellectual activity and cooperative inquiry

Identify ways to arouse pupils' interest in a problem and its
formulation

Evaluate attempts to transmit adult values and a spirit of
intellectual curiosity

Mod 26. Lesson planning

Identify ways that materials should be organized for instruction

Recognize learning activities likely to provide for maximum learning
under given conditions of (1) pupil readiness, (2) physical
layout of classroom, (3) availability of materials

Identify ways to motivate pupils to learn

Evaluate alternative paths to the same objective

Define the following terms: behavioral objectives, sequence,
materials articulation, evaluation

Mod 27. Stating behavioral objectives

Recognize generally stated goals when translated into pupil
behavior terms

Identify objectives which include (1) descriptions of behaviors,
(2) conditions under which they are to be exhibited, and
(3) criteria for judging their quality and quantity

Identify 3 major elements in statements of objectives: (1) behavior
to be produced, (2) intermediate and terminal behavior to be
taken as evidence of achievement of objective, and (3) criterion
by which terminal behavior is initiated

Discriminate between statements that constitute evidence that learning
has actually taken place and statements about behavior from which
learning can only be inferred

Differentiate between statements of observable and of inferential behaviors

Translate conceptual objectives into behavioral terms

Rephrase poorly stated behavioral objectives in correct terms

Mod 28. Use of technological aids

Identify when and how to use media, equipment, supplies, techniques

Identify ways a teacher can use himself in conjunction with materials to create learning situations not possible without collaboration with technology

Evaluate appropriateness of specific technological aids for particular purposes

Mod 29. Use of teaching materials

Identify educational materials in terms of what makes them effective with which learning problems

Identify techniques for modifying existing materials to special purposes

Identify ways to construct materials for specific applications

Evaluate appropriateness of materials

Component 5. Style. This component should predict stable patterns of teacher behavior--the general climate he maintains in his classroom

Mod 30. Discipline

Evaluate disciplinary procedures and identify probable affects

Identify and evaluate alternative forms of punishment

Judge what action is appropriate to observed infractions

Identify alternatives to common disciplinary procedures

Identify appropriate standards of conduct to be maintained

Mod 31. Classroom management skills

Estimate the probable effectiveness of a wide variety of management behaviors

Identify ways to help students understand directions

Recognize situations in which pupils can assess how social values and norms operate to control individuals' behavior

Recognize situations which can cause conflict and identify adequate responses to such situations

Mod 32. Maintenance of classroom environment

Identify as a reasonably well-structured environment for the learner one that is supportive, fairly controlling, but with a stress on self-delineation and negotiation

Define constructs such as warmth, critical thinking, openness, consciousness of cultural differences and recognize them when they occur

Judge the extent to which classroom climate is permissive, admiring, praising, accepting, self gratifying, reassuring, unthreatening, non-valuing, non-comparing, and identify reasons why

Identify ways of creating an environment in which threat to the self of the learner is reduced to a minimum and a differentiated perception of the field of experience is facilitated

Area 4. Modules in The Area of Professionalism

Modules in this area were derived in part from the Elementary Teacher Education Models. The most obvious difference between them and modules in the last areas lies in their not directly involving work with pupils. As in the last area, examples of specific item objectives are listed under those components derived from the Model Programs, but there were relatively fewer of them to be found there.

Component 1. Professional Awareness. This component has to do with the degree to which a teacher actually maintains contact with new ideas in education; how alert he is.

Mod 33. Awareness of current educational trends

Identify ways of keeping up with innovations

Demonstrate familiarity with new media, programs, materials

Demonstrate familiarity with current research in education, recent books and articles, issues under current discussion

Mod 34. Research consumption skills

Identify standards for evaluating research

Interpret statistical data on tests and in research reports

Identify findings which bear on local problems

Differentiate which research findings have practical relevance

Component 2: Self-Improvement Skills. These modules attempt to assess the resources a teacher has for evaluating, experimenting with, and changing his educational skills.

Mod 35. Observational skills

Identify which of a selected taxonomy of behaviors were exhibited
by an individual or group being observed

Discriminate teaching behaviors with sets of categories reflecting
several dimensions of teaching

Interpret interaction analysis data

Mod 36. Experimental approach to teaching

Identify ways of creating and testing out original solutions to
educational problems

Evaluate new programs and patterns of organization as they apply
to his own classroom

Identify ways of reassessing and modifying his own teaching

Derive testable hypotheses from actual classroom events, problems,
or issues

Identify criteria that can be used to test various theories

Mod 37. Ability to evaluate his own teaching objectivity

Identify ways of evaluating his own teaching

Identify ways to gauge effects of attempts to change his own behavior

Construct small studies to study his own progress toward mastery of
technical skills and strategies

Identify ways to identify the kinds of maneuvers he habitually uses

Component 3. Interpersonal Skills. Modules in this component are designed
to assess the teacher's effectiveness in working with other adults and adult
institutions and organizations to improve his effectiveness as a teacher.

Mod 38. Understanding of School-Community Problems

Identify common school practices which work to the disadvantage of certain groups, such as the overinterpretation of IQ tests

Identify and explain major social changes which place a strain on schools, such as urbanizations, youth revolt

Recognize some problems facing American education, such as the lack of sufficient reading material for the culturally disadvantaged

Mod 39. Understanding of School As An Organization

Identify levels of responsibility with the school system

Recognize how the informal organization differs from the formal one

Identify behaviors which assume a bureaucratic from ones which assume a collegial principle of authority

Differentiate between primary (manifest) and secondary (latent) functions of educational organizations

Conceptualize interaction components which characterize bureaucratic positions and thereby assign legitimacy to authority

Mod 40. Working with parents

No examples available

Mod 41. Working with peers

No examples available

Mod 42. Working with superiors

No examples available

Component 4. Professional Attitude. Because of the emphasis in the examination lies heavily on cognitive abilities and knowledges, little space is devoted to non-cognitive measurements, but some effort should be made to assess the commitment of the candidate to teaching as a career, his sense of

professional ethics, and most particularly his commitment to the idea that learning to teach is a lifelong process. The development of modules in this area, however, should be postponed until some of the more accessible means of improving the examination have been exploited.

Area 5. Personality Integration

This area is, of the five listed, the most difficult one in which to establish interpretable modules, or even to define them in measurable terms. It is assessed indirectly in many modules listed under other headings; indeed, it is doubtful whether many candidates could perform "well" on the test (on any properly designed scoring key, that is) unless they are pretty stable and acceptant of themselves. In any case, no attempt will be made here to identify the components of this important though elusive quality.

Concluding Remarks

To construct a test which would yield information on no more than the 42 modules defined here will be a formidable task. But so is the task of predicting teacher performance from a group test a formidable task. If the former were less formidable, there would be reason to doubt that its accomplishment could contribute much to the accomplishment of the second task. And the definition of the domain of teacher competence provided by these 42 modules is woefully incomplete and inadequate. Fortunately, the history of measurement has repeatedly shown that useful measurements can be obtained from the crudest of instruments--and often are. Perhaps this would be another instance!

IV. THE SHAPE OF THE NEW EXAMINATION

Besides the innovations in scoring procedures alluded to briefly in Chapter I, innovations in item format, administration, and response mode are proposed, each of which is designed to improve the quality of the information yielded. Before going into any detail about these innovations, it may be useful to provide a brief overview in the form of an inventory of the resources available for test modification. These proposed changes will be discussed under the heading of item formats, modes of administration, and response modes.

Item formats. The traditional format for items on the National Teacher Examinations is a conventional multiple-choice form, printed in a booklet. The input to the candidate is entirely verbal--that is, he must read the item to find out what the problem is. Some items in the examination tend to involve quite a bit of reading--particularly those which involve "realistic" problems in the solution of which the candidate is supposed to apply the knowledge the test is designed to assess.

One of the most powerful changes proposed for the new examination is to present the problems in the form of films or videotape recordings of actual classroom situations projected on a screen instead of in verbal terms, so that a candidate's ability to "read" behavior will become as important a factor in his test performance as his ability to read the printed word is now.

Somewhere between these two contrasting types lies another viable possibility: the problem might be presented in audio-visual mode--that is, projected on a screen in printed (or pictorial) form and simultaneously read aloud by a narrator. This would seem also to reduce the importance of

reading ability in determining a candidate's test performance, although it would not directly depend on his ability to read behavior.

There are then, three item formats: verbal, audio-visual, and situational.

Model of Administration. Introduction of the possibility of administering part or all of the test to candidates by film or closed-circuit television raises a new possibility. In a conventional test the candidate has before him all of the content of all of the items on at least one subtest simultaneously and for a substantial period of time. During this time he can read and re-read any part he likes, spending more time on some items than on others.

If the items are projected one by one--as is the case when they are presented on closed-circuit television--information appears serially; each bit is there for a while and then it is gone forever. The candidate must attend to whatever the examiner presents at the time when it is presented, and only then.

These two modes of administration will be referred to as static and dynamic, respectively.

Response Modes. The conventional best-answer item confronts a candidate with what is basically a discrimination task--that of selecting the best among four or five alternatives. There is a similar item format called the cluster true-false item which seems worthy of consideration for this application. This type of item requires the candidate to react to each of several alternatives individually as true or false. Superficially, the task it presents differs from the one presented by the best-answer only in that when the item is a cluster true-false item, the candidate does not

know how many of the alternatives are correct in a given instance; but when it is a best-answer item he knows that only one of the five is incorrect, which makes the best-answer item appear intrinsically easier.

This is true, of course, only of inferior best-answer items. A well-written item confronts the candidate with a carefully constructed set of discriminations he must make among alternatives which are neither entirely "true" nor entirely "false," but vary in degree in such a way that considerable judgment may be needed to identify the best one.

However, when knowledge rather than judgment is to be measured, the cluster true-false item seems suited to cover more ground than the best answer in a given amount of time.

A third item format is proposed for use in items designed to measure judgment, as in items where the candidate's task is to evaluate alternative strategies or solutions to a problem. To stipulate that among four or five alternative strategies one must be clearly the best is to confront the candidate with an unrealistic situation. In problems encountered in the classroom such a situation is very rare. Sometimes there is not even one solution to a problem that is better than another; often there are two or three equally good ones. Sometimes all are equally bad. Sometimes there is no consensus even among experts as to which of a set is best--one will prefer A, another B.

It is proposed that problem items be presented in a format called judgmental. In this mode, the candidate will be asked to rate each alternative on its own merits on a five-point scale, with the understanding that all alternatives can be rated equally high or equally low, or in any other pattern the candidate prefers.

These last two response modes adapt themselves particularly well to items in either situational or audio-visual format and dynamically administered, although they may be used with other formats and modes of administration as well.

To give the best idea of what changes might be made in the examination it seems expedient to leap ahead and to describe what examinations might look like which incorporated all of the changes. It would be an interesting but certainly mad idea to spring all of these changes on the National Teacher Examinations clientele at one time. (It may be interesting to speculate on the probable reaction as the following pages are read.) It would be far more prudent to try first one and then another, adopting one by one the changes which prove advisable on after empirical investigation and devising better ways of proceeding to replace those which do not work out. But as a convenient way of advancing a number of proposals simultaneously, this seems as good as any.

One other point which should be mentioned before embarking on this venture has to do with the modules upon which scoring is to be based. There is no intention of organizing the sequence in which the examination is presented in terms of modules; it is expected that (as indeed is the case in the actual classroom) the candidate should never know what knowledge or skill he will need next. Nature does not confront man with sets of problems organized by content or by the approach used in solving them, so why should the examination? The score on any given module will be based on an item here, a response there, a rating in the other place. The order of problems and items on the test should be organized in some other meaningful order such as, perhaps, the order in which they might be encountered in a teacher's day, week, or term.

Part 1: The Biographical Questionnaire

The first contact a candidate has with the examination comes not when he reports at the center, but when he reads the Bulletin of Information and fills out his Registration Form. Here is an unrivalled opportunity to collect data just as likely to be relevant to teaching success as answers to test questions, in the guise of answers to general information questions not unlike those on the present registration form, but more numerous.

The Bulletin itself, which contains the instructions for filling out the form, should begin with some set induction material explaining the general purpose and nature of the examination and how the information supplied by the candidate will be used. It should point out that, unlike many examinations which the candidate may have taken in the past, this one is not primarily designed to evaluate his knowledge and compare him with some norm group. Its purpose is to inventory his knowledges, skills, experiences, and his preferences in teacher assignments, so that his particular pattern of qualifications may be used to match him with those vacancies in the profession he would prefer and would be most likely to succeed in.

For the fee that the candidate sends in with his application he will receive a diagnostic profile indicating areas in which he is relatively strongest and weakest, with those areas (if any) in which he needs further preparation clearly identified. He will also receive a description of types of positions for which he has the best chance of being hired. If he chooses (and only if he so chooses) his name and address will be sent to school systems seeking candidates with his pattern of characteristics so that they may get in touch with him. For a modest additional fee he will be sent a list of school systems looking for teachers with his pattern of characteristics to which he may apply for employment.

The section in the Bulletin which contains instructions on how to fill out the registration form will be longer than it now is, because instead of the 15 items now contained it will include a larger number of items. Otherwise it will be similar to it. The first few items will ascertain the name, address, etc., of the candidate and the other information presently obtained. Others will refer either to the candidate's past experiences, training, and the like or to the kind of position he would like to obtain--the kinds of pupils, situation, subjects, etc., he prefers.

In the next portion of this report are listed a number of sample items typical of some types of items that might be used. Following it is a discussion of the possible uses of each.

SAMPLE QUESTIONNAIRE ITEMS

12. How much teaching experience have you had?
1. More than five years
 2. Two to five years
 3. One year (beside student teaching or internship)
 4. Student teaching or internship only
 5. None whatever (skip to Question 22)

If your answer to the last question was No. 5: None whatever, do not answer the following questions but skip ahead to Question 22. Otherwise, answer the questions in terms of the class and school you now teach in or (if you are no longer teaching) in terms of the last class and school you taught in, or the one you presently teach in.

13. Which of the following best describes this school?
1. Rural school
 2. Small, suburban
 3. Large, suburban
 4. Small town
 5. Urban

14. How did you feel about this school?
1. Liked it very much
 2. Liked it most of the time
 3. Just accepted it as a job to do
 4. Often unhappy with it
 5. Thoroughly disliked it and was glad to leave

15. How many books do you think there were in the typical student's home?
1. Several bookcases full
 2. Two bookcases of books
 3. About one bookcase of books
 4. About ten books
 5. Less than five books
16. Where was the school located?
1. Northeast (Conn., Del., Mass., Me., N.H., N.J., N.Y., Pa., R.I., Vt.)
 2. Southeast (D.C., Fla., Ga., Md., N.D., S.C., Va., W. Va.)
 3. South Central (Ala., Ark., Ky., La., Miss., Okla., Tenn. Texas)
 4. North Central (Ill., Ind., Iowa, Kans., Mich., Minn., Mo., Nebr., N.Dak., Ohio., S.Dak., Wis.)
 5. Pacific and mountain (Ariz., Calif., Colo., Idaho., Mont., N.M., Nev., Ore., Utah, Wash., Wyo.)
 6. Outside the Continental U.S.
17. Which of the following best describes the community in which the school was located?
1. Suburb in a metropolitan area of more than 2,000,000 population
 2. Suburb in a metropolitan area of 500,000 to 2,000,000
 3. Suburb in a metropolitan area of 100,000 to 500,000
 4. In a city (not a suburb) of more than 500,000
 5. In a city of 50,000 to 500,000
 6. City or town of 10,000 to 50,000
 7. Town of less than 10,000
 8. Farm, ranch or other open country
18. In comparison with other children you have known, how much freedom did parents give your students to do the things they wanted to do?
1. Much more freedom than most
 2. Somewhat more freedom
 3. The same amount of freedom
 4. Somewhat less freedom
 5. Very little freedom
19. How often were you disturbed when pupils left their work unfinished?
1. Almost always
 2. Frequently
 3. Sometimes
 4. Rarely
 5. Never

20. In comparison with other students you have known, how often did your students question what you told them about subject-matter?
1. Considerably more often than average
 2. Somewhat more often than average
 3. About an average amount
 4. Somewhat less often than average
 5. Considerably less often than average
21. How often do you think their parents encouraged your students or otherwise show interest in their school work and scholastic achievement?
1. Constantly--deeply interested and encouraged them a great deal
 2. Frequently--interested and gave them encouragement
 3. Sometimes--show occasional interest and encouragement
 4. Rarely--not particularly interested or encouraging
 5. Never--not interested or encouraging at all
22. Which of the following best describes the type of teacher you are? One who....
1. Outlines the problem in a general way, but leaves it up to the student to decide what is really needed and how to go about doing it
 2. Explains clearly what is to be done, but leaves it up to the student to decide how to do it
 3. Gives specific enough instructions so that there is little question as to what should be done and how to go about doing it
 4. Is specific as to what should be done and how and makes periodic checks to see if students are proceeding properly
 5. Is very specific as to what and how things should be done and closely supervises students and makes suggestions
23. How often have you changed your mind about your future field of work or occupation?
1. Never
 2. Only once
 3. Two or three times
 4. Four times or more
 5. Have not made any plans
24. How much voluntary reading did you do during the last year?
1. Several books every week
 2. About a book a week
 3. One to three books a month
 4. Two or three books a year
 5. One or less books a year

25. How many times during the past year or so have you gone to an evening lecture on some serious topic (other than required lectures)?
1. Not at all
 2. Once or twice
 3. Three or four times
 4. Five or more times
26. How many books do you yourself own (not including textbooks for your present courses, but counting serious paperbacks)?
1. Less than ten
 2. Ten to 30
 3. 31 to 75
 4. More than 75
27. At what level would you prefer to teach?
1. Nursery school, kindergarten, preschool Go to question 00
 2. Primary grades (1-3) Go to question 00
 3. Intermediate (4-6) Go to question 00
 4. Junior high school (7-9) Go to question 28
 5. Senior high school (10-12) Go to question 28
 6. Junior college (13-14) Go to question 00

Answer the following questions only if you marked 4 or 5 above.

28. In what kind of school would you prefer to teach?
1. Public school
 2. Private, nonreligious, nonmilitary
 3. Protestant denominational
 4. Catholic
 5. Jewish
 6. Military
 7. Does not matter
29. With how large a graduating class?
1. Less than 50
 2. 50 to 99
 3. 100 to 199
 4. 200 to 299
 5. 300 to 399
 6. 400 to 599
 7. 600 to 799
 8. 800 to 1000
 9. More than 1000
 10. Does not matter
30. With what proportion of the class going on to college (including junior college)?
1. Less than one-fourth
 2. From one-fourth to one-half
 3. From one-half to three-fourths
 4. More than three-fourths
 5. Does not matter

31. Of the subjects listed below, which one would you prefer to teach? (Mark only one.)
1. Art
 2. English (including speech and literature)
 3. Foreign language(s)
 4. Mathematics
 5. Music
 6. Physical education
 7. Sciences (Physics, biology, etc.)
 8. Shop or commercial (e.g., typing courses)
 9. Social sciences (history, civics, etc.)
59. Which subject would you like to teach the least? Use the alternatives in the preceding question. (Mark only one.)
Go to question 62.
62. As far as you personally are concerned, which one of the requirements below is the most important in any job or profession you would consider going into?
1. Opportunity to use my special abilities and talents
 2. Prospects of an above-average income
 3. Freedom to be creative and original
 4. Opportunity to work with people rather than with things
 5. Opportunity to be helpful to others and/or useful to society in general
 6. Stable, secure future
 7. Compatibility with the kinds of people with whom I would be working
 8. Avoidance of work under relatively high pressure
 9. Relative freedom from supervision by others
63. For women only: Fifteen years from now would you like to be:
1. A housewife with no children
 2. A housewife with one or more children
 3. An unmarried career woman
 4. A married career woman without children
 5. A married career woman with children
 6. Right now I am not certain
64. How do you evaluate your ability to hold students' attention and to present opinions and thoughts in a clear and orderly way?
1. Excellent
 2. Good
 3. Average
 4. Poor
 5. Very poor

65. How do you feel about giving a speech before a group of other teachers?
1. Dislike it very much
 2. Dislike it somewhat
 3. Do not particularly mind it
 4. Enjoy it somewhat
 5. Enjoy it very much
66. Which one of the following do you think is closest to describing your personality?
1. Difficult to really get to know
 2. Have a few really close friends and a number of acquaintances
 3. Friendly and easy-going; have a lot of friends
 4. Very jolly; the "life-of-the-party" type

Rank the following types of high school subjects in terms of how much you liked them. (On the answer sheet mark choice 1 for most liked, choice 2 for next most liked, and so on until you mark choice 5 for least liked.)

67. Physical sciences (such as chemistry, physics, and mathematics)
 68. Natural sciences (such as biology)
 69. Social sciences (such as history, current events, civics and government)
 70. Literature and creative writing
 71. If you are a boy, courses such as shop and auto mechanics, if you are a girl, courses such as home economics
72. How much do you agree or disagree with the statement: "A man can be well informed even if there are many subjects upon which he does not have a definite opinion."
1. Strongly disagree
 2. Disagree
 3. Neutral
 4. Agree
 5. Strongly agree

Discussion of Biographical Questionnaire Items

Samples of five types of questionnaire items which might be used are shown above. Items 12 and 27 function primarily as branching items which route the candidate around items irrelevant to his situation. Item 12 shunts inexperienced teachers past questions about teaching experiences; item 27 directs the candidate to questions relevant to the type of position he prefers.

Items 11 and 27 can perform a second function which is classificatory in nature, a function also performed by items 28 to 31, inclusive. Candidates can be sorted according to amount of experience and type of position preferred so that the computer can search for teachers meeting certain specifications in this area.

Items 13, 16, and 17 are designed to produce what might be called objective descriptions of past experience. Responses to these items could also be used for classificatory purposes, but are primarily intended for another use in conjunction with the type of item described next, that is items yielding subjective descriptions of experience.

Subjective descriptions are obtained from items like numbers 14 and 15 and 18 through 22. Here a candidate is asked to infer or guess something or to express an opinion. Such items seem to show some potential for getting at attitudinal elements related to modules 19, 30, 32, 35, 36, 38-- toward teaching, pupils, the community, discipline, etc. Such inferences might be enhanced by relating opinions to "objective" descriptions. At the simplest level, responses to item 14 indicating how well the candidate liked the school he taught on could be related to the location and size of the school (items 16 and 17) to infer what kind of school the candidate likes

(or dislikes), and further checked against the kind of school he prefers as described on items 28-30. These cross-checks could be used to extract more valid information than might be obtained if only the conventional attitude items were used.

Finally, a few items--numbers 23 to 26 and 62 to 72--are included which are best described simply as personality items. This is an area where a sure and gentle touch is needed, particularly in a context of selection; but there seems to be no other way of getting at the teacher's self-concept and other matters related to the fifth area of competence. Perhaps items like numbers 23 to 26 or 62 and 63, which sound factual will be more useful than opinion items like numbers 64 to 66.

In any case, a questionnaire like this, "administered" as part of the registration process and therefore using no testing time at all, filled out on a mark sensing sheet (such as an NCR form, perhaps) so that it can be processed economically and rapidly, seems to offer a particularly rich source of highly useful information which ETS could obtain and dispense more efficiently than anyone else.

The Paper-and-Pencil Section

On the surface, the paper-and-pencil section of the new examination would look very much like the present examination. Items designed to measure the first seven modules would look very much like those on the present Common Examinations. The fact that the items would have been selected for high internal consistency within modules--to produce, if possible, bimodal distributions of scores on each module--might prevent the generally bright candidate from finding the entire examination rather easy, as he must do now. There should be some items that any candidate (except a modern Renaissance man!) would perceive as difficult.

The insertion of items to measure modules 11 to 13 should further this impression. To get a clear idea of what is meant by items likely to produce a bimodal distribution on a module about which a candidate is ignorant, the typical white middle-class reader need only attempt the following items:

A BLACK GHETTO CULTURE MODULE (No. 11)

73. Whom did "Stagger Lee" kill in the famous blue legend?
1. His mother
 2. Frankie
 3. Johnny
 4. His girl friend
 5. Billy
74. If a man is called a "blood" then he is a
1. Fighter
 2. Mexican American
 3. Negro
 4. Hungry Hemophile
 5. Redman or Indian
75. If you throw the dice and seven is showing what is facing down?
1. Seven
 2. Snake eyes
 3. Boxcars
 4. Little Joes
 5. Eleven
76. In "C.C. Rider" what does "C.C." stand for?
1. Civil Service
 2. Church Council
 3. Country Circuit (Preacher)
 4. Country Club
 5. Cheatin Charlie (the "Boxer Gunsel")
77. Cheap "chitlings" (not the kind you purchase at a frozen-food counter) will taste rubbery unless they are cooked long enough. How soon can you quit cooking them to eat and enjoy them?
1. Fifteen minutes
 2. Eight hours
 3. Twenty-four hours
 4. One week (on a low flame)
 5. One hour

78. Hattie Mae Johnson is on the county. She has four children and her husband is now in jail for nonsupport, as he was unemployed and was not able to give her any money. Her welfare check is now \$286 per month. Last night she went out with the biggest player in town. If she got pregnant, then nine months from now, how much more will her welfare check be?
1. \$80
 2. \$2
 3. \$35
 4. \$150
 5. \$100
79. The "Hully Gully" came from
1. East Oakland
 2. Fillmore
 3. Wants
 4. Harlem
 5. Motor City
80. Many people say that "Juneteenth" (June 19) should be made a legal holiday because this was the day when
1. The slaves were freed in the U.S.
 2. The slaves were freed in Texas
 3. The slaves were freed in Jamaica
 4. The slaves were freed in California
 5. Martin Luther King was born
 6. Booker T. Washington died
81. Jazz pianist Ahmad Jamal took an Arabic name after becoming famous. Previously he had some fame with what he called his "slave name." What was his previous name?
1. Willie Lee Jackson
 2. LeRoi Jones
 3. Wilbur McDougal
 4. Fritz Jones
 5. Andy Johnson
82. A "gas head" is a person who has a
1. Fast-moving car
 2. Stable of "lace"
 3. "Process"
 4. Habit of stealing cars
 5. Long jail record for arson

The typical college graduate seldom encounters a test which contains a subtest made up of items so completely baffling as these. Even the non-mathematician finds one or two mathematics items he can make a stab at on the mathematics test, because such tests are composed of items selected

to discriminate at various levels of ability. Such a policy would be inefficient on a modular test on which few items could be assigned to any one module. The mathematics knowledge module should have mathematics items as opaque as those above to anyone not well-grounded in the subject.

The items listed above boldly violate at least two widely honored principles of test construction--that items should constitute a representative sample of the content area to be measured, and that they should call for knowledges that have intrinsic importance. The items probably call for trivial bits of information, and they certainly do not sample the full range and scope of black ghetto culture--much of which is shared with white middle-class culture.

If a group of mathematicians set out to build a 12-item mathematics test which would effectively separate mathematicians from other educated men, this is exactly the kind of module they would need to construct; one asking for esoteric bits of knowledge that only mathematicians are likely to know.

If items from all modules were intermingled--instead of being segregated as is the common practice--the chances are that the modules would function even more effectively, and the generally well-informed candidate would not be overwhelmed by encountering 12 such frustrating items in a row. The poorly-informed one might; but he should be accustomed to it. In other words, both should do about as well as they do now overall, but instead of passing about an equal number of items in each area they would tend to pass all in some areas and none in others.

The paper-and-pencil section would also contain items related to modules in the areas of teaching skill and professional knowledge (scattered among the others). These, too, would need to be somewhat esoteric to fulfill their function. The face validity of the examination might appear to be jeopardized by this. In a sense, this part of the examination would have less face validity if the type of problem item which appears on the present Common Examinations were replaced by such apparently trivial factual items.

This would be more than offset by the strong face validity of the film test administered the same day. Examinees would quickly realize that ability to deal with teaching problems was being measured in this way, and that the paper-and-pencil items were designed to measure technical knowledge. As a matter of fact, it might be a good thing for the image of the teaching profession if candidates ignorant--and contemptuous--of the content of professional courses encountered some of these items and realized that there is something to be learned about professional education!

It has been suggested that a module like the one above might be used to measure interest in an area as well as knowledge of it. Only someone really interested in a topic is likely to know the fine points of it. Teachers who say on the biographical questionnaire that they would like to teach in the inner city, but who fall down on module 11, may be suspected either of deceiving themselves or of trying to deceive someone else.

The same argument leads to the possibility that a high score on a module related to a component of teaching skill or professional knowledge might identify someone not only well versed in but strongly committed to that aspect of teaching or knowledge, and likely to perform competently in it.

The Film Section

Although videotape equipment would probably be used at least in the development, and quite likely also in the production, of this portion of the new examination, the test will probably have to be administered by way of 16 mm. sound film. At the present level of development of closed-circuit television hardware, it would be too hazardous--and expensive--to depend on installations of playback equipment in all the administration centers. Moreover, it is much easier and more economical to produce virtually identical multiple copies of one film than of one videotape.

The candidate taking this portion of the examination would have before him the usual SCRIBE answer sheet on which to indicate his answers.

The questions would be projected on a screen in a semi-darkened room, beginning with special orientation and instructions and practice exercises of sufficient duration to allow any necessary adaptation to the light level to take place.

The instructions would make it clear that the candidates will be shown films of actual classroom episodes. After each episode, a number of statements referring to the episode will be presented one at a time each bearing an item number, and the candidate's task will be to mark the appropriate space after that number on his answer sheet to indicate whether he agrees with the statement (marks it TRUE), disagrees (marks it FALSE), or is undecided (marks ? or omits to mark it at all). The test is structured as one of ability to perceive events in the classroom and make judgments about their meanings. It is also made clear that the judgments will have to be made under some time pressure--that ability to "think on one's feet" will be more important than ability to reach decisions after careful deliberation.

In reading the examples below, it should be remembered that at any given moment the candidate has before him only what is printed between two horizontal lines. Only one statement is on the screen at one time, and no statement appears simultaneously with the episode, although a portion of the action may reappear superimposed over the statement being evaluated in some instances.

After each statement the number of the module it refers to is indicated. (This would not be shown in the actual test, of course.)

EXAMPLES OF FILM ITEMS

SITUATION: The film shows a teacher discussing a picture on the bulletin board showing some smiling people standing in front of a 747 jet airliner with a second-grade class. At first the teacher discusses what is going on in the picture, and then what is going to happen next; finally she develops the point that these people are going on a trip.

Statement 83: The main concept the teacher was trying to develop was a little bit too difficult for most of the pupils to grasp. (Mod 18)

Statement 84: The teacher failed to ask any questions calling for divergent production. (Mod 15)

Statement 85: The picture functioned effectively as a means of arousing pupils' curiosity and interest. (Picture reappears on screen.) (Mod 29)

Statement 86: A somewhat tighter control (and a bit less calling out) would have produced a better environment for learning. (Mod 32)

Statement 87: If the teacher had interrupted the class to call for quiet, it would have taken even longer to develop her main point. (Mod 30)

SITUATION: The film shows a teacher at a chalkboard talking about a diagram of a fish. The class of teenagers is passively attending for the most part, although two girls near the back seem to be whispering or giggling every time they appear on the screen.

Statement 88: The teacher asked at least one question which challenged the students to make a hypothesis. (Mod 15)

Statement 89: As soon as the teacher noticed the two girls whispering he should have directed a question to one of them. (Film shows girls whispering.) (Mod 30)

Statement 90: The girl at the end of the front row was more interested in the lesson than most of the students were. (Film shows girl.) (Mod 17)

Statement 91: The teacher's explanation of the way fish eggs are fertilized was a procedural one. (Mod 16)

Statement 92: The drawing on the board showed too much irrelevant detail. (Film shows drawing.) (Mod 29)

Construction of a test of this type would be simplified to where it should be fully practicable. There would need to be on the ETS staff a production team whose know-how was primarily technical. When this crew was sent out to collect raw tapes their instructions would be relatively easy to follow, calling for good quality recordings of typical behaviors in classes of certain grades, while certain subjects were being taught, or possibly during seat work, supervised study, small group work, etc. It would not be necessary for them to look for certain kinds of events or the like.

This operation would use the existing console, the remote pan and tilt gear, two cameras, and one videotape recorder, plus minor accessories. All of this equipment has been acquired already.

A second team whose expertise is in teacher education and item construction would view the raw tape, pull out certain episodes, and then make up alternatives according to specifications based on the modules, using as many episodes as necessary. Outside consultants might or might not be used in this phase.

The technical team would then take over and put together a test with video, audio, titles, and special effects as indicated, under the supervision of the professional team. The final test would then be transferred to film for quantitative reproduction.

For this phase some new equipment would need to be procured. A Syntron Generator (or its equivalent) to merge titles and behavior recordings and a second VTR with editing capability would be minimal (a third VTR and camera and a second remote control unit should be added before long).

Staffing for this operation should not be too difficult--technical and professional skill and know-how need not be combined in any one person, save perhaps a chief editor who should understand the capabilities and limitations of the equipment. Equivalent forms should be no more difficult to develop than they now are, since the exact content of the situational clips is not crucial.

The Teaching Problems Section

One of the most appealing things about a teaching test based on filmed episodes is the promise it offers of being able to confront a candidate with realistic teaching problems to solve. The reader will have noticed that the Film Test just described does not do this. It essentially taps a candidate's ability to see and understand what is going on in the classroom.

The fact is that it is very difficult to construct film problem items because it is difficult to find or produce films which structure such problems clearly. So for the present it is recommended that the attempt be abandoned, and that problem items be presented verbally, in an audio-visual mode, and that the candidate be asked to respond to them in a dynamic mode--that is, under time pressure.

This test should follow the film test in time, since it involves the same set-up of projector, answer sheet, etc. The instructions for this test will be to respond by marking option 1, 2, 3, or 4 on each item instead of True or False. The instructions might read somewhat as follows:

This next test you will take will differ from most tests you have taken before in that there will not always or even usually be one correct answer to any item. Each item will consist of three parts: a situation, a problem, and a number of suggestions as to how to solve the problem. You will be asked to evaluate each suggestion in turn on its own merits, and then to indicate your evaluation by marking the appropriate space after the number on your answer sheet corresponding to that suggestion, as follows:

If the suggestion is excellent--if you would follow it immediately, mark space 1.

If the suggestion is pretty good--would solve the problem--but you would try to think of something better if you had time, mark space 2.

If the suggestion is fair--if it provides only a temporary but not a real solution, mark space 3.

If the suggestion is poor--if it does not solve the problem or would do actual harm--so you would not do it, mark space 4.

If you cannot make up your mind in the time allowed, you may leave the item blank or mark space 5.

To make this clearer, consider the following perhaps rather trivial example:

SITUATION: A mosquito lands on your arm and prepares to sting you.

PROBLEM: What should you do to keep from getting stung?

Suggestion A. Swat it.

This is an excellent idea--it solves the problem permanently. You would mark space 1.

Suggestion B. Go get an aerosol can and spray the porch with it.

This is a pretty good idea, but an awful lot of trouble. Mark space 2.

Suggestion C. Shoo it away.

This is fair at best--it offers you time to think by providing a temporary solution, but the bug will almost certainly return before long. Mark space 3.

Suggestion D. Blow cigarette smoke at it.

This is another fair suggestion--better than the last, perhaps, but still deserves to be marked in space 3.

Suggestion E. Ignore him; maybe he will go away.

This is a poor suggestion; if you followed it you would almost certainly get stung. Mark space 4.

Before you begin the test, you should know that you are probably going to find out that you will not usually have as much time to make up your mind as you might like to have; this is intentional. This test is designed to let you demonstrate how well you can think on your feet--how capable you are of making quick and accurate decisions. We are not interested in how well you can figure out what to do if you have plenty of time to think things over. So do your best!

SAMPLE PROBLEM ITEMS

In reading the following items, it should be borne in mind that, like those on the film test, only what appears between a pair of horizontal lines will be available to the candidate any one point in time. A verbal sketch of a problem situation will be presented first, visually and audibly. Then a problem will be stated. Finally, one suggestion will appear at a time, together with a brief rating guide as follows:

1. excellent
2. good
3. fair
4. poor

Thus Suggestion 94 would appear as follows:

94. Laugh along with the class, and then go on with the discussion.

1. excellent
 2. good
 3. fair
 4. poor
-

For brevity, the rating guide is omitted in each of the examples that follow.

SITUATION: You have been discussing the story of Damon and Pythias with your sixth-grade class, using it as an example of masculine friendship. One of the boys (who has already established himself as a trouble maker) makes a negative comment about the relationship in which the word queen plays a prominent part, and the class laughs. This is the first incident of this type in your new class.

PROBLEM: What should you do next?

Suggestion 93. Disregard the remark and ask a question which will attract and hold the pupils' interest.

Suggestion 94. Laugh along with the class, and then go on with the discussion.

Suggestion 95. Capitalize on the remark by discussing this aspect of the story in a matter-of-fact way.

Suggestion 96. Take this opportunity to find out and allay any curiosity your pupils may have about homosexuality.

Suggestion 97. Indicate your disapproval firmly but without making an issue of it, and be sure to speak to the boy alone later.

Suggestion 98. Nip the situation in the bud by sharply reprimanding the boy and requiring him to report to you after school.

SITUATION: At the close of the first day of school you sit down to plan the seating in your seventh grade class, planning to assign seats at least for the first few days in alphabetical order so you can learn the pupils' names more easily. However, among your pupils you have noticed three or four who look like foreigners and are not quite as neatly dressed and groomed as the rest. One of these is a lively, restless boy with curly black hair and (you note) a marked body odor.

PROBLEM: Where would you put him?

Suggestion 99. In a group with the other pupils of his social class.

Suggestion 100. In the front row near your desk.

Suggestion 101. Near an open window.

Suggestion 102. Near one or two very popular boys.

Suggestion 103. Wherever he comes in the alphabet.

SITUATION: When you attend the first orientation meetings for new teachers at your first school, you find that there are a number of traditional procedures all teachers are expected to follow such as ways of lining pupils up to move them through the halls, a certain way pupils are supposed to arrange test and homework papers, etc. Many of these are in direct conflict with what you have learned about pupil-teacher relationships and incompatible with procedures you have developed in your pre-service experiences and which have worked well for you.

PROBLEM: Evaluate each of the following suggestions

Suggestion 104. Go along with the recommended practices.

Suggestion 105. Try to get some of the other teachers to work with you on replacing these practices with more up-to-date ones.

Suggestion 106. Speak to your supervisors about changing their practices.

Suggestion 107. Ask for a transfer to a more congenial school.

Suggestion 108. Do things your own way in your own classroom, but conform to the rules when you are outside.

SITUATION: In talking to some of the experienced teachers who are to be your colleagues in your first job in a ghetto school, you are given the following bits of advice by one or another of them on how to establish good working relationships with your first class.

PROBLEM: Evaluate each suggestion on its own merits, regardless of the way it is expressed.

Suggestion 109. Be firm, demanding at first; it will be okay if you ease off later in the term.

Suggestion 110. If you have any management problems send for help; they never do a thing in the office anyway and will probably be happy to have something to do.

Suggestion 111. Try to iron out your own problems in your own room; you will be better off in the long run.

Suggestion 112. Be understanding of the faults of the students, you must realize the environment they have to cope with.

Suggestion 113. The only way to teach these kids anything is to find out something that interests them, and go on from there.

SITUATION: After a four-day unit on solving equations with one unknown, you give a test and your class does so poorly as to indicate that little or no learning has taken place. You feel bad because you had planned for each day a careful explanation with examples and illustrations, of one of the three main techniques you want them to use, with a careful review on the fourth; and had given them problems to do at home each night besides.

PROBLEM: If you had a chance to do the unit over, what would you do differently?

Suggestion 114. Not try to cover so much ground so that you could explain each point more clearly.

Suggestion 115. Study the homework papers each night and spend more time on problems the pupils had trouble with.

Suggestion 116. Have the students try to solve some problems first, and then show them the easiest way.

Suggestion 117. Try to motivate the pupils better before you start.

Suggestion 118. Give short daily mastery tests and do not go ahead until preceding content has been mastered.

Suggestion 119. Arrange for the pupils to work some problems in class when you can give individual help.

SITUATION: A fifth-grade girl seems not to have any friends nor to speak up often in class, even though she is well-dressed, pleasant looking, and cooperative.

PROBLEM: How could you help this child?

Suggestion 120. Take the first opportunity to speak to her personally and try to find out what her problem is.

Suggestion 121. Call on her regularly during class discussions.

Suggestion 122. Find out some special skill or talent she has and have her demonstrate it to the class.

Suggestion 123. Talk to her after school and tell her to be more outgoing and to volunteer more often in class.

Suggestion 124. Get in touch with the school psychologist and arrange for him to see her.

Discussion of Teaching Problems Test

No attempt has been made to tie these Problem Items directly to the modules proposed above. This reflects a conflict between the nature of the modules and that of the items. The modules, being derived from performance goals of teacher education programs, are prescriptive in nature. That is to say, each "criterion" was constructed to yield a certain behavior considered "correct" or "right." The items, on the other hand, do not assume that a correct solution can be identified for each problem. On the contrary the items are sufficiently vague and lacking in details so that a candidate's response must reflect his own attitudes, beliefs, assumptions almost as much as any knowledge he may have that is relevant to the problem. This, it must be admitted, is an attempt to make a virtue of necessity. A problem stated with enough detail so that one and only one solution to the problem can be defended as correct (or best) must either be so easy that anyone with good horse sense can recognize the solution, or it must be far too time-consuming to be useful on a test of reasonable length. But when an item is presented in rather broad terms, the candidate himself must fill in the details according to his own background and predispositions. The candidate reading the problem about Damon and Pythias (see page 62) who sees the communication of information as the primary function of education will almost certainly look at the suggestions differently than one who conceives the function of education as that of preparing students to cope with the world we all live in today.

About all the modules are good for is to indicate areas which problems should represent. Suggested solutions should perhaps be proposed by a team of educators representing a variety of viewpoints about teaching, and a

varied selection of them included in the test. This would ensure a variety of options for the candidates; but it would be very difficult to say in advance what such items would measure.

If the proposal for "tailored scoring" made in the chapter is adopted, the client could be invited to take the test himself, indicating how he would want his candidates to rate each suggestion. The computer could then search current files for candidates who marked the test that way, and a list supplied to the client.

A factor analysis of the test might be made to develop some meaningful dimensions along which candidates could be ordered and in terms of which feedback could be given to candidates about themselves.

If these items do not fit the specifications proposed in this report, why are they included? Because, next to the film items, they seem to possess the highest face validity of any yet uncovered. This apparent validity seems to come, not so much from the realism of the problems, as from the nature of the suggested solutions. The fact that, instead of being right or wrong they represent the kind of thing one might actually try, makes them appear to call for professional judgment (rather than more intelligence).

BY WAY OF CONCLUSION: SOME PARTHIAN SHAFTS

While we are not at all certain what combination of events makes a good lesson or what combination of qualities makes a good teacher, the potentially better teacher is one who is able to plan and control his professional behavior--to teach many kinds of lessons, to reach many diverse learners, to create different social climates, and to adopt a wide range of teaching strategies to constantly changing conditions.... Our definition of the "good" teacher is not someone who teaches in a certain way but someone with the capacity to create and carry out strategies and maneuvers that he modifies constantly in response to student behavior.

Joyce and Harootounian (1967, pp. 94 and 112) have here stated the definition of teacher competence on which any future teacher examination must be based if it is to meet the measurement needs of the future well enough to survive. Such a definition is, of course, more palatable to the researcher than to the tester, and more palatable to the tester than to the educator. In the past, ETS has tried to run with the educator and to operate as though someone knew "what combination of qualities makes a good teacher." (namely, us). In the future, ETS might better walk with the researchers. Only in this way can the teacher examination program move forward; only if all claims to knowledge of what makes a good teacher are abjured can the examinations be useful to the schools in finding better teachers, and only then can the day approach when the nature of teacher effectiveness is clearly understood and when ETS can provide a valid test of teacher competence.

The reason why the schoolman does not like Joyce and Harootounian's definition is that it is useless to him. The school administrator must act as though he knew what makes a good teacher, even if he doesn't. Someone has to decide which teacher to hear, which to promote, which to fire. That someone is not ETS. ETS can assist in the process by giving the decision-maker as much of the information he needs about the teachers from whom he

must choose as we possibly can. ETS can also help him study the effects of his decisions and gradually improve them. This is the proper function of the teacher examination service.

Neither of these functions requires ETS to know what makes a good teacher. A test or test battery which will predict how well a teacher will teach is out of reach; it is neither necessary nor possible to construct such a test at present. But a test or test battery that will predict how a teacher will teach may be possible, and would certainly be useful.

The bulk of this report has been devoted to some suggestions as to how such a test might be constructed. Chapter II contains a sketch of a domain of behaviors all of which may be or have been defended as likely to contribute to success in teaching. Chapter III presents a somewhat more structured list of behaviors that have been identified as important enough to qualify as behavioral objectives in modern programs for training teachers. Chapter IV proposes certain specific techniques for measuring--or at least for predicting--the degree to which a candidate will behave in these ways.

The biographical items administered at the time of registration for the examination are meant to make available information about the candidate's past experiences and future plans which may cast light on his future performance.

The paper-and-pencil test, under the assumption that the cognitive knowledge a candidate has is a factor in determining his behavior, is designed to assess that knowledge in the most efficient way possible.

The film test is designed to assess the candidate's functional knowledge--that is, his ability to relate theory, research, and his own experience to situations he may encounter on the job.

The teaching problems section is designed to elicit responses from the candidate which will indicate how he is likely to cope with problems in the classroom--which will measure professional judgment.

All of these suggestions involve innovation in a greater or lesser degree, but the principal change that needs to be made in the examinations is discussed in Chapter I. The new examination will hardly be any better than the old unless there is a drastic change in the way the examinations are scored and interpreted. There are three steps in the evaluation process as it applies to human behavior: securing a behavior sample, quantifying the behaviors, and evaluating them. The upper limits on the quality of the evaluation obtained is determined by the first two steps--the steps that are the proper concern of the measurement service. Both are important; there is room for substantial improvement in both steps; some suggestions have been made with regard to both which seem feasible and likely to succeed. Whether the full potential of the new examination is realized or not depends ultimately on the third step--on how the test performances are evaluated and used by the school and college administrators themselves. This is not under ETS control; all we can do is remove limitations resulting from test content and scoring procedures, and then use the gentle art of persuasion.

References

- Allen, D. W., & Cooper, J. M. A proposed new program for elementary teacher education at the University of Massachusetts. Appendix I Final Report, October 31, 1968, University of Massachusetts, Contract No. OEC-O-8-089023-3312(010), United States Department of Health, Education, and Welfare.
- Allen, D. W., Cooper, J. M. A proposed new program for elementary teacher education at the University of Massachusetts. Appendix II Final Report, October 31, 1968, University of Massachusetts, Contract No. OEC-O-8-089023-3312(010), United States Department of Health, Education, and Welfare.
- Allen, D. W., & Cooper, J. M. A proposed new program for elementary teacher education at the University of Massachusetts. Final Report, October 31, 1968, University of Massachusetts, Contract No. OEC-O-8-089023-3312(010), United States Department of Health, Education, and Welfare.
- Bestor, A. Educational wastelands. Urbana, Illinois, Illinois University Press, 1953.
- Bowley, R. L. Teaching without tears. London: T. J. Winterson, 1961.
- Broudy, H. S. Research and the dogma of behavioral objectives. Urbana, Illinois: University of Illinois. Unpublished manuscript. (Mimeographed)
- Bruner, J. S. The process of education. Cambridge, Massachusetts: Harvard University Press, 1960.
- DeVault, M. V. Wisconsin elementary teacher education project. Volume I Element specifications. Madison, Wisconsin: University of Wisconsin School of Education, 1969.
- DeVault, M. V. Wisconsin elementary teacher education project. Volume II Element specifications. Madison, Wisconsin: University of Wisconsin School of Education, 1969.
- DeVault, M. V. Wisconsin elementary teacher education project. Volume III Element specifications. Madison, Wisconsin: University of Wisconsin School of Education, 1969.

- Dickson, G. E. Educational specifications for a comprehensive elementary teacher education program. Final Report, October, 1968, University of Toledo, Project No. RFP OE-68-4, Contract No. OEC-O-8-089026-3310(010), United States Department of Health, Education, and Welfare. 2 Volumes.
- Fuller, Frances. Concerns of teachers: A developmental conceptualization. American Educational Research Journal, 1969, 6, 207-226.
- Gorman, C. J. The University of Pittsburgh Model Program. Project No. 8-9020, Contract No. OEC-O-8-089020-3309(010), October 1968, United States Department of Health, Education, and Welfare, Office of Education, Bureau of Research.
- Johnson, C. E., Shearron, G. F., & Stauffer, A. J. Georgia education model specifications for the preparation of elementary teachers. Final Report, October, 1968, University of Georgia, Project No. 8-9024, Grant No. OEC-O-8-089024-3311(010), United States Department of Health, Education, and Welfare.
- Joyce, B. R. The teacher--innovator: A program to prepare teachers. Final Report, October, 1968, Teachers College, Columbia University, Project No. 8-0019, Contract No. OEC-O-8-089019-5307(010), United States Department of Health, Education, and Welfare.
- Joyce, B. R., & Harootounian, B. The structure of teaching. Chicago, Illinois: Science Research Associates, 1967.
- Justiz, T. B. A method for identifying the effective teacher. Paper presented to the American Research Association Annual Meeting, Los Angeles, February 5-8, 1969.
- Michigan State University. Behavioral science elementary teacher education program. Final report, October 31, 1968, Project No. 8-9025, Contract No. OEC-O-8-089025-3314(010), United States Department of Health, Education, and Welfare, Volume II.

- Morse, W. C., & Wingo, G. M. Psychology and Teaching (3rd Edition).
Glenview, Illinois: Scott, Foresman & Company, 1969.
- Raths, L. E., Harman, M., & Simon, S. B. Values and Teaching. Columbus,
Ohio: Charles E. Merrill Books, Inc., 1966.
- Rivlin, H. N., & Robinson, Valda. The preparation of urban-teachers:
A syllabus. Working paper One, 1968, Fordham University, the NDEA
National Institute for Advanced Study in Teaching Disadvantaged Youth.
- Schalock, H. D. Northwest Regional Educational Laboratory Program.
Project No. 8-9022, Contract No. OEC-O-8-089022-3318(010), October
1968, United States Department of Health, Education, and Welfare,
Office of Education, Bureau of Research.
- Sorenson, A. G., & Gross, Cecily F. Teacher appraisal: A matching process
Unpublished manuscript.
- Sorenson, A. G., Husek, T. R., & Yu, Constance. Divergent concepts of
teacher role: An approach to the measurement of teacher effectiveness.
Journal of Educational Psychology, 1963, 54, 287-294.
- Sowards, G. W. The Florida State University Model Program. Project No.
8-9021, Contract No. OEC-O-8-089021-3308(010), October, 1968,
United States Department of Health, Education, and Welfare, Office of
Education, Bureau of Research.
- Stoller, Nathan. Preliminary thoughts on possible bases for an assessment
instrument in teaching. Hunter College, 1970. Unpublished manuscript.
Syracuse University. Specification for a comprehensive undergraduate and
inservice teacher education program for elementary teachers. Final
Report, October, 1968, Project No. 8-9018, Contract No. OEC-O-8-0918-
3313(010), United States Department of Health, Education, and Welfare.

Wittrock, M. C. The evaluation of instruction: Cause and effect relations in naturalistic data. University of California at Los Angeles Evaluation Comment, Center for the Study of Evaluation, May 1969, Vol. 1, No. 4.
(mimeographed)

APPENDIX

COGNITIVE FACTORS IN TEACHING STYLE

Donald M. Medley
Educational Testing Service
Princeton, N. J.

Russell A. Hill
Research for Better Schools
Philadelphia, Pa.

A paper presented at the Annual Meeting
of the American Educational Research Association
Minneapolis, Minnesota, March 4, 1970

COGNITIVE FACTORS IN TEACHING STYLE

The findings that I am going to describe to you today were a byproduct of a larger study involving 70 first-year intern teachers in a large metropolitan area in the eastern United States. These 70 teachers were visited in their classrooms four times each by a pair of trained observers, and their behavior was recorded for not quite half an hour on each visit. Fifty-three of the 70 teachers had also taken the Common Examinations of the National Teacher Examinations just before beginning their first year of teaching, and we were able to retrieve their answer sheets for use in the present analysis. These 53 secondary school teachers, on whom both behavior records and test data were available, constitute the subjects with whom we are concerned today. The group included teachers of all four major subjects--science, mathematics, English, and social studies--at both the junior and senior high school level.

One of the two observers who visited each teacher was trained in the system of Interaction Analysis developed by Flanders (Amidon & Flanders, 1963) and recorded verbal behavior according to that system. The other was trained to use a different technique, OScAR 4V (Medley, Impelletteri, & Smith, 1966), and recorded the same verbal behaviors using that system.

All observations were intercorrelated and submitted to a principal components analysis on the basis of which 15 scoring keys were built, 8 for OScAR, and 7 for the Flanders' system (Medley, & Hill, 1968, 1969). Scores on these 15 keys accounted for about two-thirds of all of the variance in the observations. These scores constituted the measures of teacher behavior, or style, used in the present study.

The form of the Common Examinations of the National Teacher Examinations taken by the 53 teachers contained 345 multiple-choice items. The 345 items were written according to a table of specifications which called for items representing 19 different content areas. About half of the items were designed to sample the teacher's knowledge of subject-matter content commonly included in secondary school curricula, such as science, mathematics, English, etc. The other half were designed to measure knowledge of the content of professional education courses--history and philosophy of education, teaching principles and practices, and so on. Subscores

were obtained for each teacher on each of 19 "subtests" made up of items from one of the 19 content areas.

Answer sheets were available for 91 teachers, including the 53 who were observed plus 38 others enrolled in the same program but not observed in their classrooms. These 91 papers were submitted to an analysis of variance of the form originally suggested by Hoyt (1941), to study the internal structure of the test. Hoyt used a two-way design without replication, items by candidates; in this case, there were 90 degrees of freedom for candidates, 344 for items, and 30,960 for error. We extended Hoyt's design by partitioning the 344 degrees of freedom for items into two portions. One portion, with 18 degrees of freedom, estimated between-items variance from variation between items on different subtests only; the other, with 326 degrees of freedom, estimated the same variance by comparing only items on the same subtest.

If you will consult Table 1, you will note that the sum of squares for error was also partitioned in an analogous fashion. One portion, with 1,620 degrees of freedom, estimated errors of measurement from interaction between candidates and items on different subtests, and the other, with 29,340 degrees of freedom, estimated errors of measurement from interaction between candidates and items on the same subtest.

The fact that the former mean square is larger than the latter indicate that σ^2_{cs} is greater than zero; that is, that there is an interaction between candidates and subtests. Therefore we may not assume that all of the 19 subtests are measuring the same function since some candidates tend to do better on one subtest than they do on others. Or, to put it differently, since the rank order of true scores of the 91 candidates varies from one subtest to another, we must conclude that the subtests measure different functions. The practical interpretation is that there is information in the subtest scores that does not appear in the total score, so we must retain at least some subtest scores for further analysis.

The question remains: how many, and which subtest scores should we retain? To answer this question we partitioned the 18 degrees of freedom between subtests, and the 1,620 for candidates by subtest interaction, into 18 parts each. In the case of subtests, there was one degree of freedom

for each part, in the case of interaction, there was 90 degrees of freedom for each portion.

In making these partitions we used 18 orthogonal contrasts among subtests, reflecting 18 a priori hypotheses about how the content areas sampled by the 19 subtests might differ. The null hypothesis was rejected in eight instances and accepted in ten as regards between-subtests variation. The null hypothesis was rejected in nine instances and accepted in nine as regards interaction. Table 2 presents a condensed version of the analysis of variance in which all non-significant mean squares have been pooled with their respective error terms.

In order to conserve all of the information in the test scores it was necessary to retain scores on the 11 subtests shown in Table 3. Table 3 also shows the contrasts found to be significant, and the mean reliability per item of each subtest.

This last statistic is, of course, equivalent to what should be obtained by using the Spearman-Brown formula backwards on each subtest, "prophesying" the reliability of a one-item test in each instance (Gulliksen, 1950, pp. 77-79). The mean reliability per item gives a pretty good idea of the extent to which each subtest is saturated with its own principal component, and its magnitude is independent of the number of items on the subtest.

The analysis of primary interest to us today is the one summarized in Table 4. Each of the 15 behavior dimensions in turn was regressed on the 11 NTE subtests. Eight of the equations obtained are shown in the table. Neither the multiple correlations nor any of the beta weights in any of the other seven equations was significantly different from zero, so none of them are shown.

Since only two of the 15 equations resulted in a multiple correlation whose probability under the null hypothesis was less than .05, and since only nine of the 165 beta weights met this criterion, these findings should be regarded as tentative only. Because data of this type are so rare, however, they may be worth peeking at. Attempts to predict teacher competence (as measured by various criteria) from teachers' scores on cognitive tests have been uniformly unsuccessful in the past (cf. Barr, 1948). Here we have asked a different question. Instead of trying to predict some

Table 1

Analysis of the Performance of the
91 Teachers on 19 Subtests of the
Common Examinations of the National Teacher Examinations

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	
			Obtained	Expected
Candidates	90	253.20	2.81	$345\sigma_c^2 + \sigma^2$
Subtests	18	348.64	19.37	$91\bar{K}\theta_s^2 + 91\sigma_i^2 + \sigma^2$
Items (in subtests)	326	1,202.18	3.69	$91\sigma_i^2 + \sigma^2$
Interaction, Candidate by Subtest	1,620	528.49	0.33	$\bar{K}\sigma_{cs}^2 + \sigma^2$
Residual	<u>29,340</u>	<u>5,089.91</u>	0.17	σ^2
Total Variation	31,394	7,422.41		

$$\bar{K} = \frac{1}{18} \left(345 - \frac{\sum K_i^2}{345} \right)$$

K_i = number of items on subtest i

$i = 1, 2, \dots, 19$

Table 2

Pooled Analysis of Variance
of Scores of 91 Teachers on 345 Items of the
Common Examinations of the National Teacher Examinations

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares
Candidates	90	253.20	2.81
General Knowledge (vs. Professional)	1	96.30	96.30
Science and Mathematics (vs. rest of General Knowledge)	1	17.07	17.07
Science (vs. Mathematics)	1	18.46	18.46
English (vs. Social Studies and Fine Arts)	1	45.91	45.91
Literature (vs. English)	1	35.46	35.46
Social Studies (vs. Fine Arts)	1	16.32	16.32
Foundations (vs. Teaching Principles and Practices)	1	27.61	27.61
History and Philosophy (vs. School and Society)	1	48.84	48.84
Items (within Subsets)	336	1,244.85	3.71
Candidate x General (vs. Professional Knowledge)	90	57.29	0.64
Candidate x Science and Mathematics (vs. Other General Knowledge)	90	78.17	0.87
Candidate x Science (vs. Mathematics)	90	42.62	0.47
Candidate x English (vs. Social Studies and Fine Arts)	90	41.15	0.46
Candidate x Literature (vs. English Mechanics and Effectiveness)	90	53.44	0.59
Candidate x English A (vs. English B)	90	23.00	0.26
Candidate x Social Studies (vs. Fine Arts)	90	44.42	0.49
Candidate x Foundations (vs. Teaching Principles and Practices)	90	20.39	0.23
Candidate x History and Philosophy (vs. School and Society)	90	22.18	0.25
Pooled Residual	30,150	5,235.73	0.17
Totals	31,394	7,422.41	

Table 3

Subtests Scores on the Common Examinations and
Contrasts for which the Null Hypothesis Was Rejected

Subtest	Number of Items	Mean Reliability Per Item	Contrasts
Science	30	.16	+ + + 0 0 0 0 0 0
Mathematics	20	.20	+ + - 0 0 0 0 0 0
English A	30	.11	+ - 0 + - + 0 0 0
English B	25	.09	+ - 0 + - - 0 0 0
Literature	19	.21	+ - 0 + + 0 0 0 0
Social Studies	30	.08	+ - 0 - 0 0 + 0 0
Fine Arts	16	.18	+ - 0 - 0 0 - 0 0
Teaching Principles and Practices	60	.03	- 0 0 0 0 0 0 0 - 0
History and Philosophy of Education	20	.03	- 0 0 0 0 0 0 0 + +
School and Society	22	.02	- 0 0 0 0 0 0 0 + -
Psychological Foundations plus Teacher Role plus School Organization	73	.04	- 0 0 0 0 0 0 0 + 0
Total	345		

amorphous construct called "competence" we have tried to predict stable patterns of classroom behavior which may be regarded as elements of teacher style, and which are clearly defined in operational terms.

The most impressive finding in Table 4 is the multiple correlation of .66 between Lecturing Behavior (as scored on the Flanders' records) and performance on the NTE. Inspection of the beta weights in the equation indicates that the scores a teacher obtains on science items and on items related to teaching principles and practices are principally responsible for this relationship. Teachers who do better on the science items lecture more; teachers who do better on the teaching principles and practices items lecture less.

Results obtained in the larger study indicate that science teachers as a group tend to lecture more than other teachers, so the contribution of the science subtest to the regression equation may be a function of subject taught, in part at least. However, since there were only seven science teachers among the 53 included in the study, it is likely that teachers of other subjects who had high science subtest scores also tended to act like science teachers no matter what subject they taught.

The negative relationship between lecturing and knowledge of items related to teaching principles and practices is intriguing, suggesting as it does that the teacher who lectures may do so only because he does not know any better way to teach!

In looking at the rest of the results in Table 4 let us remember that they are only suggestive, not conclusive. And in doing so let us ask ourselves the question: if a teacher gets his highest score on items of one particular type, what kind of teaching behavior would you expect him to exhibit in his classroom?

The teacher scoring highest on science items tends to be high on Lecturing Behavior, as we have noted, and also on the Modified Content Cross and the I-D Contrast on Student Responses.

What these dimensions all have in common is a sensitivity to two of Flanders' ten categories: Lecturing and Asking Questions, plus a negative weighting on pupil responses, particularly those to which the teacher reacts in a direct fashion. The teacher whose forte is science spends a lot of time dealing with subject-matter, and tends to dominate the discussion himself.

Table 4

Regressions of Teacher Behaviors
on National Teacher Examinations Common Examinations

Dependent Variables	Multi- ple R	Beta Weights on Test Subscores									
		Science	Mathe- matics	English A	Liter- ature	Social Studies	Fine Arts	Teaching Principles	History and Phi- losphy	School and Society	Psy- chology etc.
1. Lecturing Behavior (FIAT)	.66*	+.38*	+.13	+.18	-.37	-.15	+.16	-.42*	+.30	+.02	-.04
2. Lecturing Behavior (OSCAR)	.56	+.14	+.05	+.19	-.30	+.30	+.20	-.42*	-.05	+.15	-.25
3. Modified Content Cross (FIAT)	.60*	+.39*	-.21	+.31	+.05	-.01	+.17	-.12	+.25	-.14	+.02
4. Criticizing Behavior (FIAT)	.55	-.02	-.12	-.48*	+.19	-.30	-.16	-.09	+.33	+.38	+.11
5. Rebuking Behavior (OSCAR)	.48	-.06	-.18	-.27	+.26	-.52*	+.07	+.01	+.47	+.33	+.02
6. I-D Contrast on Student Response (FIAT)	.53	+.36	-.28	+.16	+.19	-.11	+.08	-.03	+.42*	+.00	-.18
7. Questioning Style (OSCAR)	.53	-.04	+.27	-.19	-.33	+.20	-.08	-.27	-.09	+.46*	-.12
8. Listening Behavior (OSCAR)	.48	-.19	+.25	-.05	-.46*	+.36	+.07	+.21	-.09	-.26	+.10

*P < .05

No patterns emerge for teachers with high scores on Mathematics or English A. Those who score highest on English B, which was designed to measure effectiveness of expression, are low on Listening Behavior. This means that pupil comments in their classroom tend to be brief; their students do not deliver monologues or speak at length without teacher interruption.

The teacher who scores highest on literature items tends to behave in the opposite fashion. He listens to his pupils more, and lets them speak at length; he seldom finds cause to rebuke them, and his own comments tend to be brief.

Teachers scoring high on social studies and art items show no clear-cut pattern; as far as subject-matter content goes, then, it appears to be science, effectiveness of expression, and literary acquaintance that relate to teaching style.

When we turn to professional knowledge, we find that (as noted) teachers who know most about teaching principles and practices tend to teach by question-and-answer rather than by the lecture method. This is confirmed on both of our Lecturing Behavior measures.

The teacher who scores highest on the 20 items devoted to the history and philosophy of education also scores high on Rebuking Behavior and on the I-D Contrast on Student Response. Such a teacher rebukes pupils frequently (but without rancor or hostility), and at the same time reacts positively to pupil responses to teacher questions. The suggestion is that such a teacher's classroom has a noisy but positive climate.

The teacher who scores high on "School and Society" items is one who asks questions calling for thoughtful and original answers, and who asks students to evaluate and elaborate their own responses. Since this teacher is also the one who criticizes or rebukes pupils most frequently, this suggests a teacher who challenges pupils with difficult questions.

A more sophisticated analysis of these data--perhaps one using canonical correlations--might have made them appear more clear-cut, but we were reluctant to base any more complex analysis than the one reported on so slight a data base lest we inflate their apparent importance more than we may already have done. Let us conclude by stating two inferences which we feel the data do justify.

First, they strongly suggest that the amount and kind of cognitive equipment a teacher possesses is an important element in determining his teaching style.

Second, there is considerable promise in the methodological strategy used in this study--that is, in trying to relate teacher knowledge to teacher behavior. If we had 100 teachers or more--instead of 53--the results in Table 4 suggest that we would have learned quite a bit about how to predict teacher behavior from tests administered to them before they began to teach, and that we might learn something about the dynamics and etiology of teaching styles as well.

References

- Amidon, E. J., & Flanders, N. The role of the teacher in the classrooms. Minneapolis, Minnesota: Paul S. Amidon and Associates, Inc., 1963
- In E. J. Amidon & J. B. Hough (Eds.), Interaction Analysis: Theory, Research, and Application. Reading, Massachusetts: Addison-Wesley, 1967.
- Barr, A. S. The measurement and prediction of teaching efficiency: A summary of investigations. Journal of Experimental Education, 1948, 16, 203-283.
- Gulliksen, H. Theory of mental tests. New York: Wiley, 1950.
- Hoyt, C. J. Test reliability estimated by analyses of variance. Psychometrika, 1941, 6, 153-160.
- Medley, D. M., & Hill, R. A. A comparison of two techniques for analyzing classroom behaviors. American Educational Research Association, Chicago, Illinois, 1968. In R. C. Burkhardt (Ed.), The Assessment Revolution: New Viewpoints for Teacher Evaluation. Albany, New York State Education Department, Division of Teacher Education and Certification, 1969.